

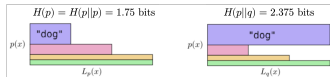
Introduction to Machine Learning

Source Coding and Cross-Entropy



Learning goals

- Know connection between source coding and (cross-)entropy
- Know that the entropy of the source distribution is the lower bound for the average code length



SOURCE CODING AND CROSS-ENTROPY

Cross-entropy is the average length of communicating an event from one distribution with the optimal code for another distribution (assume they have the same domain \mathcal{X} as in KL).

$$H(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{1}{q(x)} \right) = - \sum_{x \in \mathcal{X}} p(x) \log (q(x))$$

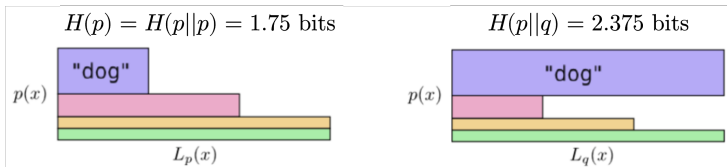
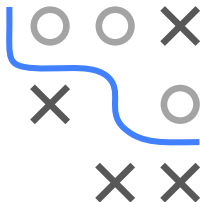
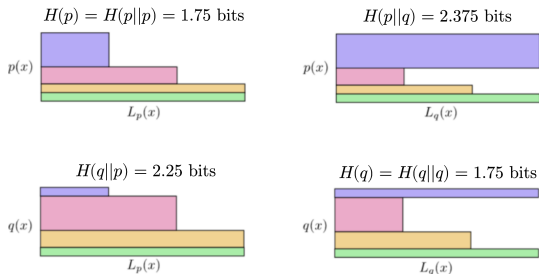


Figure: $L_p(x)$, $L_q(x)$ are the optimal code lengths for $p(x)$ and $q(x)$

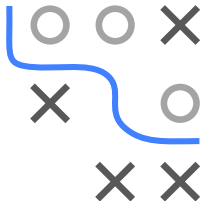
We directly see: cross-entropy of p with itself is entropy:

$$H(p||p) = H(p).$$

SOURCE CODING AND CROSS-ENTROPY



Credit: Chris Olah



- In top, $H(p||q)$ is greater than $H(p)$ primarily because the blue event that is very likely under p has a very long codeword in q .
- Same, in bottom, for pink when we go from q to p .
- Note that $H(p||q) \neq H(q||p)$.

